On the Deployment of Large Scale NSaaS

Thomas Deiß, Bertold Dickhaus, Dereje Kifle
Nokia
Ulm, Germany

***Abstract*—Network Slicing is an enabling technology of 5G mobile networks. Multiple logical networks, satisfying different service level requirements can be deployed on the same infrastructure. Each of these logical networks can be described as a network service descriptor as defined by ETSI NFV. Although ETSI NFV defines support for physical network functions and for deployments across multiple datacenters, network services are taken to a new level of scale in 5G network slices. In this paper we describe issues caused by this level of scale and which have to be solved when describing and deploying a 5G network slice as a network service (NSaaS). We also describe how these issues have been tackled in the 5G-TRANSFORMER project.**

Keywords—network slicing, orchestration

I. Introduction

5G has been developed to support a wide range of requirements, e.g. peak bandwidth up to 10Gbps, latency as small as 1ms, and support for huge number of devices, without requiring that all requirements hold simultaneously for a single service. The requirements can be satisfied by different deployments of network functions and corresponding configurations. E.g. for low latency, a user plane function (UPF) can be deployed close to the mobile edge, whereas for enhanced mobile broadband it is better to place UPFs in data centers to achieve statistical multiplexing gains. Each of these different deployments forms a logical network, named a network slice instance (NSI). 5G has built-in support for network slicing both in the radio access network (RAN), especially at the air interface, and in the core network (CN).

Each NSI consists of a set of network functions in the CN, expected to be virtualized network functions (VNF), and it uses physical network functions (PNF) – radio, distributed, and central units, (RU, DU, CU) – of the radio access network (RAN). Part of the network functions in the RAN may be virtualized as well, nevertheless the RAN will consist to a large extent of PNFs. ETSI NFV has defined network services and their corresponding descriptors (NSD) [1]. A network service is a set of VNFs and PNFs, connected by virtual links (VL). The actual deployment is done by one or multiple orchestrators. Multiple orchestrators can be arranged in a hierarchical manner, allowing to use specialized orchestrators for areas such as the transport or radio networks.

Here, we highlight issues on deploying network slices as a service (NSaaS) on large geographical scale, which are due to the atypical nature of the network services representing 5G NSIs. Examples are V2X networks along highways or at intersections, communication services along railway lines, or mobile broadband communication in megacities.

This work has been partially funded by the EC H2020 5G-TRANSFORMER Project (grant no. 761536).
Corresponding author: thomas.deiss@nokia.com

II. Background

Network slices defined by 3GPP for 5G networks [1] are logical networks, including both RAN and CN. NSIs may differ regarding the features or optimizations they support. Also, NSIs may have different SLAs, providing connectivity with different characteristics such as latency to the UEs. These differences in features and SLA requirements imply that network functions may be deployed differently, e.g. in edge clouds or in a central datacenter. An NSI consists of smaller parts, named network slice subnet instances (NSSI). E.g., a 5G NSI may consist of NSSIs for the RAN, the CN, and transport.

ETSI NFV has defined NSDs [1] and an interface for the lifecycle management of network services [4]. Network services are sets of VNF or PNFs, connected via VLs. Connection points (CP) are used internally to denote interface of VNFs and service access points (SAP) define the external interface of network services. Network services can also be composed hierarchically from other network services.

III. Issues

A. PNF Orchestration

It is state of the art to develop CN functions as VNFs and not any more as dedicated appliances. This in mind, the 5G CN has been specified by 3GPP as a service-based architecture, targeting a deployment as VNFs. RAN is different. Although some parts of gNbs such as the central units (CU) can be virtualized, the remaining parts are expected – distributed units (DU) – or will definitely – radio units/antenna (RU) – be deployed as dedicated appliances. Instead of creating a suitable amount of VNF instances and the connections among them, a RAN orchestrator has to discover the topology of deployed gNbs, determine the subset of them needed for the NSI to be deployed, and eventually (re)configure these gNbs such that they can be shared among multiple NSIs.

The set of deployed gNbs is changing throughout the lifetime of a NSI, as new gNBs will be installed, existing ones will be upgraded or taken out of service. Although a network service may include PNFs and as such could include gNbs, this is not feasible in our case, as the NSD would have to be adapted permanently to the changes of the physical infrastructure. Also, the selection of the subset of gNbs sufficient for a specific NSI cannot be captured in an NSD.

Therefore, we describe the RAN as an abstract network service. An example is depicted in Figure 1. This network service connects an SAP for the air interface with an SAP towards the CN. UEs may connect to the air interface SAP (SAP_NR). The UEs themselves are not considered a part of

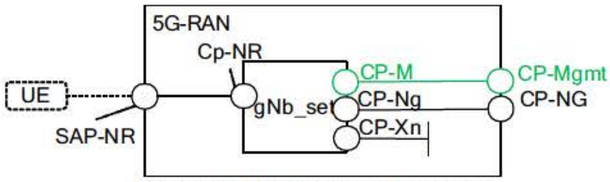

Figure 3: RAN network service

the NSI. Additionally, the network service provides an SAP for management. In this example, we kept the virtual network for handovers within the network service. It would also be possible to use the network for NG for handovers.

B. *Air Interface SAP*

To use an SAP for NR as shown in Figure 1, one has to indicate that this is an SAP for NR, and provide air interface related information. As SAPs are derived from CPs, we can use the `layerProtocol` and `cpProtocol` attributes of a CP descriptor [5] to indicate that this is a NR SAP.

Air interface related information is on the one hand the coverage area and on the other hand it describes UE properties to be supported. Coverage area could be described as a geographical area as in geographical information systems. The `SapData` information element has been extended to include this information at instantiation time. An important property of UEs are the frequencies on which they connect to the RAN. UE frequency is one of the capabilities defined by 3GPP for UEs [6]. We extend the SAP definition with an attribute `mandatoryCapabilities` to describe capabilities of UEs, including the frequency bands, that the RAN or a specific NSI must support. Another attribute `unneededCapabilities` can be used to explicitly describe those capabilities that the RAN or NSI does not have to support and which could be omitted to optimize the deployment regarding OPEX, CAPEX, or energy consumption.

C. *Degree of Distribution*

The NSI is distributed across a large geographical area. A multitude of data centers would be used, especially in combination with multi-access edge computing (MEC) [2]. A single SAP may be mapped to multiple links on transport level.

D. *Multiplicity*

When deploying a network service with multiple VNFs across multiple datacenters one can create instances in different ways. On the one hand, an orchestrator could instantiate a network service instance per data center. Alternatively, the orchestrator can instantiate one instance of the network service with its VNF instances deployed in many datacenters. Deciding whether to create a single network service instance or multiple ones is a challenge for orchestrators. This challenge gets even more complex if there are functions common to several NSIs, e.g. a common authorization and authentication service. Such a common function might exist with one or a few instances, but would be connected to many other VNF instances in different datacenters.

E. *Network Slice Coverage*

Availability of NSIs in the RAN is defined with tracking areas (TA) granularity. However, an NSI may be unsupported in some cells of a TA due to factors such as RAN incapability or unavailability of shared or dedicated resources, etc.

F. *Inter slice connections*

The 5G CN consists of slice specific functions, such as the UPF, as well as slice common functions such as the Network Slice Selection Function (NSSF) [1], which is actually the logical function being aware of the different NSIs in a 5G network. A dedicated NSI can be expected to contain slice specific functions only. To make use of the slice common functions, this specific slice and the common functions have to be connected. One may consider the slice common functions to be deployed in an NSI for the common functions. This distinction allows to manage these parts separately, and even by different operators.

The dedicated NSI has to be connected with the common NSI. This introduces another level of complexity as gateways among the NSIs are needed, agreements on authorization and authentication mechanisms need to made, multiple management frameworks might be in use etc. On the other hand, these clearly defined interfaces provide a clear separation of concerns, also allowing a place where SLA requirements can be monitored and enforced, e.g. protecting the common NSI from a misbehaving dedicated NSI. Note, the common NSI would connect to multiple dedicated NSIs.

IV. POTENTIAL SOLUTIONS

The 5G-TRANSFORMER [7] system [8] will be able to handle information on the coverage area at instantiation time as part of the `SapData` information element. This information can be used by the service orchestrator (5GT-SO) and passed to the mobile transport and computing platform (5G-MTP). The 5G-MTP can have plugins to handle different types of infrastructure, e.g. with a domain-specific orchestrator for radio networks. Such a radio orchestrator can do the actual deployment of the RAN. This solution allows the service orchestrator to be independent of specific radio knowledge.

REFERENCES

[1] ETSI GS NFV-IFA 014, V3.1.1, Management and Orchestration, Network Service Templates Specification, 2018.

[2] http://www.etsi.org/mec

[3] 3GPP TS 23.501 V15.4.0, System Architecture for the 5G System; Rel. 15, 2018.

[4] ETSI GS NFV-IFA 013 V3.1.1, Os-Ma-Nfvo reference point - Interface and Information Model Specification, 2018.

[5] ETSI GS NFV-IFA 011 V3.1.1, Management and Orchestration; VNF Descriptor and Packaging Specification, 2018.

[6] 3GPP TS 38.306 V15.4.0, User Equipment (UE) radio access capabilities, Rel. 15, 2018.

[7] http://5g-transformer.eu

[8] A. de la Oliva et. al., 5G-TRANSFORMER: Slicing and Orchestrating Transport Networks for Industry Verticals, IEEE Communications Magazine, August 2018.